

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sebagaimana diketahui, dalam beberapa tahun terakhir berbagai teknologi telah mengalami perkembangan yang signifikan, diantaranya ialah teknologi informasi dan bioteknologi. Bioteknologi berasal dari ilmu di bidang biologi, dan teknologi informasi berhubungan dengan bidang informatika. Perpaduan antara ilmu biologi dan matematika merupakan dua bidang ilmu yang sangat berbeda, akan tetapi ilmu pengetahuan terkini menunjukkan perkembangan yang sebaliknya sehingga menghasilkan suatu cabang pengetahuan baru yang bernama bioinformatika. Bioinformatika berawal dari peningkatan jumlah data observasi biologi sebagai bentuk gabungan antara teknologi informasi dan bioteknologi yang disebabkan oleh kemajuan bioteknologi (Nugroho, 2006).

Fatchiyah (2009) menyatakan bahwa bioinformatika dapat dikatakan sebagai ilmu yang mempelajari penerapan teknik komputasi untuk menganalisis informasi-informasi biologi. Contoh penerapan bidang bioinformatika yaitu mengelola informasi-informasi data biologi, melakukan prediksi pada bentuk struktur protein maupun struktur *Ribonucleic Acid* (RNA), analisis ekspresi gen, dan lain-lain. Aplikasi bidang bioinformatika diantaranya perakitan segmen ratusan pasang basa DNA menjadi satu sekuen genom utuh dengan ukuran mencapai milyaran pasang basa (*genome assembling*), prediksi keberadaan gen dan perangkat pelengkapannya seperti promotor dalam genom (*gene annotation*), hingga prediksi fungsinya (*functional genomics*).

Meningkatnya pemahaman manusia dalam bidang genomik tidak terlepas dari perkembangan biologi molekular modern dan jaringan internet. Ketersediaan *database* dasar (DNA, protein) yang bersifat terbuka/gratis merupakan peluang besar untuk menggali informasi berharga. Ledakan data/informasi biologi tersebut mendorong lahirnya bioinformatika (Witarto, 2003). Kumpulan data yang telah didapatkan tersebut kemudian akan bertambah banyak dan memiliki skala

yang besar. Informasi yang semakin meluap tersebut memunculkan metode yang diharapkan dapat menyelesaikan permasalahan dengan jumlah data yang besar. Maka, disini dibutuhkan solusi untuk mengatasi ledakan informasi tersebut, salah satunya yaitu dengan *Machine Learning*. *Machine Learning* merupakan salah satu disiplin ilmu untuk mengubah data menjadi informasi dari data berskala besar menjadi data yang lebih mudah dipahami. Hampir sama dengan kinerja otak manusia, *machine learning* juga membutuhkan pembelajaran agar dapat bekerja sesuai dengan kehendak. Pada bidang kedokteran, *machine learning* digunakan untuk mencari jawaban cara mendeteksi penyakit seseorang dari gejala-gejala yang ada, mendeteksi apakah seseorang mengidap penyakit jantung dari rekaman elektrokardiogram, atau mencari tahu gen yang terlibat pada penyakit kanker. Misalnya hubungan struktur dan fungsi suatu protein, hubungan antara penyakit dan profil ekspresi gen seorang pasien, atau proses timbulnya suatu penyakit (Tawakal, 2014). Pada data ekspresi gen berdimensi tinggi, dibutuhkan metode untuk dapat melakukan analisis informasi pada data tersebut, salah satunya adalah metode *Support Vector Machine* (SVM). SVM digunakan untuk menemukan *hyperplane* (pemisah) terbaik yang berfungsi memisahkan dua buah *class* atau lebih pada *input space* (Nugroho, 2003).

Pada awalnya, SVM digunakan untuk menyelesaikan masalah klasifikasi linear, namun seiring berjalannya waktu kemudian dikembangkan dengan memasukkan konsep *kernel trick* pada ruang kerja berdimensi tinggi agar dapat bekerja pada klasifikasi non-linear. SVM adalah salah satu teknik yang relatif baru dan dapat digunakan untuk memprediksi serta memiliki performansi yang lebih baik di berbagai bidang (Munawarah, 2016). SVM juga telah berhasil diterapkan pada beragam aplikasi biologi molekuler atau bioinformatika, salah satu penerapannya yaitu pada bidang pertanian untuk mendapatkan varietas unggul dari suatu tanaman.

Salah satu komoditas penting yang ditanam untuk menghasilkan tanaman bernilai mutu tinggi adalah jagung. Jagung (*Zea mays L.*) merupakan komoditas pangan yang memiliki potensi besar untuk kepentingan industri pakan dan pangan. Selain untuk konsumsi manusia, jagung juga dimanfaatkan sebagai pakan

ternak unggas. Jagung merupakan komoditas yang berpeluang untuk dikembangkan karena mempunyai peran strategis sebagai sumber utama karbohidrat dan protein setelah beras. Berbagai macam keperluan dapat dimanfaatkan pada hampir semua bagian tanaman jagung. Pada bagian tanaman jagung, batang dan daun yang masih muda dapat digunakan sebagai makanan ternak, tanaman yang telah dipanen dapat digunakan untuk pembuatan pakan atau pupuk organik (Bunyamin *et al.*, 2013).

Jagung memiliki kandungan gizi dan vitamin yaitu 355 kalori, 9.2 gr protein, 3.9 gr lemak, 73.7 gr karbohidrat, dan 10 mg kalsium. Meskipun tingkat penggunaannya belum berkembang seperti saat ini, akan tetapi permintaan jagung akan terus bertambah. Jagung memiliki kandungan karbohidrat yang tinggi dan dapat dijadikan bahan baku yang baik untuk produksi bioetanol. Oleh karena itu, jagung mempunyai kegunaan yang luas baik pangan, pakan dan energi (Bantacut, 2010 dikutip dalam Sain, 2016).

Mengingat bahwa jagung adalah salah satu tanaman pertanian yang paling penting di seluruh dunia, maka dilakukan penelitian yang memiliki implikasi yang kuat untuk perbaikan sifat genetik tanaman dan dapat memberikan keamanan sistem pangan dunia, yaitu dengan teknik *Nested Association Mapping* (NAM). NAM adalah teknik yang dirancang untuk mengidentifikasi dan membedah arsitektur genetik dari ciri-ciri kompleks dalam jagung (*Zea mays*). Jagung memiliki keragaman genetik lebih banyak daripada sistem genetika model lainnya, sehingga atribut jagung menjadi sistem yang sangat baik untuk mempelajari berbagai fenomena biologis. Keragaman ini dapat digunakan untuk mengatasi masalah mulai dari perbaikan tanaman hingga terurainya mekanisme dalam pengembangan tanaman, biokimia, dan fisiologi hingga pemahaman tentang arsitektur genetik dari sifat-sifat kompleks. NAM memiliki potensi yang luar biasa untuk penelitian sifat agronomi pada tanaman jagung dan spesies lainnya (Yu *et al.*, 2007).

Perbaikan sifat tanaman perlu dilakukan sebagai upaya untuk meningkatkan keragaman genetik. Keragaman yang luas merupakan modal dasar pemuliaan tanaman. Proses seleksi secara efektif dapat dilakukan untuk memberikan peluang

yang lebih besar dalam memperoleh karakter-karakter yang diinginkan (Sobir, 2007 dikutip dalam Sain, 2016).

Berdasarkan uraian yang telah dijelaskan di atas, maka ingin diaplikasikan metode SVM untuk mengklasifikasikan data ekspresi gen *Maize* NAM berdasarkan karakteristik organ yang telah diperoleh.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah pada penelitian ini adalah sebagai berikut.

1. Bagaimana aplikasi metode SVM pada data ekspresi gen *Maize* NAM?
2. Bagaimana pengklasifikasian terhadap data uji dengan menggunakan metode SVM ?
3. Bagaimana melakukan evaluasi terhadap hasil klasifikasi dengan menggunakan parameter akurasi metode SVM ?

1.3 Batasan Masalah

Agar penelitian tidak menyimpang dari pembahasan maka perlu diberikan suatu pembatasan masalah yaitu sebagai berikut.

1. Pada penelitian ini digunakan data ekspresi gen *Maize* NAM.
2. *Software* yang digunakan dalam penelitian ini adalah *RStudio* dengan menggunakan beberapa *package* bioinformatika dari *bioconductor*.
3. Klasifikasi SVM yang dilakukan adalah klasifikasi linear.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut.

1. Melakukan penerapan metode SVM pada data ekspresi gen *Maize* NAM.
2. Melakukan pengklasifikasian terhadap data uji dengan menggunakan metode SVM.
3. Melakukan evaluasi terhadap hasil klasifikasi dengan menggunakan parameter akurasi metode SVM.

1.5 Manfaat Penelitian

Adapun manfaat yang diharapkan dari penelitian ini adalah sebagai berikut.

1. Dapat digunakan oleh berbagai kalangan terutama disiplin ilmu bioinformatika termasuk ilmu-ilmu hayati dalam mengetahui pengaruh dari nama-nama gen yang memiliki bobot tertinggi.
2. Mampu memberikan informasi kepada berbagai kalangan khususnya disiplin ilmu biologi dan statistika terkait pengklasifikasian dan dapat memprediksi kelompok dari data yang baru.